PROPOSAL TITLE
Data Management Plan

Products of Research · As detailed in the Project Narrative document, products include:

- Simulation implementation code written in open source software **DETAILS**
- **DETAILS** datasets, file sizes on the order of **DETAILS**
- Baseline ML models for **DETAILS**.
- Open access software, tutorials, and documentation associated with each dataset.
- **OTHER**
- Open access educational materials to execute our **DETAILS** outreach plan.

How the data will be acquired and processed · The majority of the data generated for this proposal is high quality simulation data (computationally intensive to acquire). Finite Element Analysis (FEA) simulations will be run in open source software FEniCS [1] **DETAILS**. All data will be acquired by the PIs and students assigned to this project. The data will be processed with in-house scripts written in open source software. We will make all data processing scripts available through GitHub. We do not anticipate the need for any proprietary software for accessing or interpreting data.

Standards for data and metadata format and content. We will store and disseminate our data primarily through the OpenBU institutional repository. As illustrated in Fig. 1, the OpenBU institutional repository provides a permanent link to the data through the handle system, and the web interface allows the convenient display of relevant metadata. The OpenBU web interface also allows links to relevant manuscripts and code repositories. For every dataset, we will link to a public GitHub repository that contains all of the code that we create to generate the simulation data and/or analyze experimentally derived data. Our GitHub repository contains heavily commented code complete with a detailed “README.md” file. We will also provide multiple tutorials where code is organized as a Jupyter Notebook and explanations are given as additional video tutorials posted to YouTube. For every relevant manuscript submission, we will also submit a pre-print version of our manuscript and a link to the public GitHub repository to the website “paperswithcode” for maximum transparency and reproducibility of our code.

In the scenario that our storage needs exceed what is possible through OpenBU, we will work with the experienced staff of OpenBU to determine a suitable alternative. At present, our plan is to publish our data with the Dryad Repository. If publishing our data through Dryad is necessary, we will publish each dataset with similar metadata to what is illustrated in Fig. 1. We are very conscious that the accessibility of our dataset is critical to its broad adoption and ultimate impact. We have included the Mechanical MNIST example in Fig. 1 because it is representative of how we will store and disseminate the datasets proposed here.

Dissemination, Access and Sharing of Data · As illustrated in Fig. 1, we will share and preserve our data primarily through the OpenBU Institutional Repository. In the Project Narrative, we describe two different types of data that we will generate as a part of this proposal. The example dataset from Fig. 1, Mechanical MNIST, is an example of **DETAILS**. In the scenario that our storage needs exceed what is possible through OpenBU, which is possible for the larger datasets, we will work with the experienced staff of OpenBU to determine a suitable alternative. At present, our plan is to publish our data with the Dryad Repository. If publishing our data through Dryad is necessary, we will publish each dataset with similar metadata to what is illustrated in Fig. 1. We note that we have recently prepared and shared a dataset through Dryad in conjunction with our recent publication [2]. Based on the anticipated size of our datasets, we expect that our baseline machine learning (ML) models will take hours to days to train on a single core, and substantially less when training is conducted on the Shared Computing Cluster in parallel. Therefore, we believe that sharing the code we write and our data is sufficient. If our ML models take much longer to train, we will follow best practices for model sharing **DETAILS**.

Reuse and Redistribution · We do not anticipate the need for any proprietary software in accessing or interpreting our data. For the software that we create, we will post it in a public GitHub repository with a MIT License. For the relevant data that we generate it, we will share it under a Creative Commons Attribution-ShareAlike 4.0 License. Our data will be citable at a permanent link created following the handle system. In Fig. 1 we show an example of this. To cite our software, we will request that users cite the related published manuscripts.

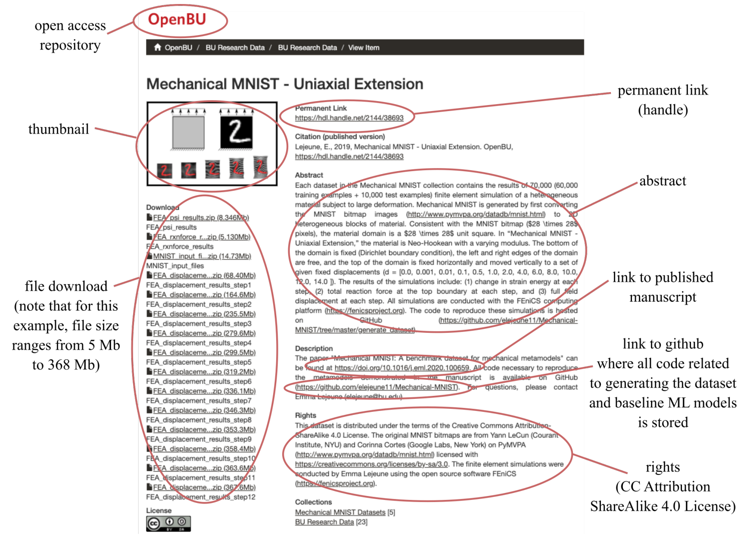

Figure 1: Illustration of the key components of OpenBU.

Archiving of Data and Timeframe for Preservation · While conducting our research, in particular prior to data curation and dissemination, we will use our Buy-In Storage at the Boston University Shared Computing Cluster and multiple external hard drives for fast access and data transfer on our laptop computers and storage redundancy. Once our curated datasets are completed, we will transfer the responsibility for sharing and preservation to the OpenBU digital repository (and potentially the Dryad digital repository). We will maintain a complete backup of each dataset in its original form for a minimum of five years after initial deposit into the repository. This will give us the flexibility to augment and verify aspects of the published dataset as needed. For every published manuscript that is funded through this proposal, we will publish all data and software concurrently with manuscript submission. We will also publish data and software in conjunction with our timeline for completing each Objective. The datasets created during this research proposal will be useful as **DETAILS**, and the software used to generate these datasets will be useful to other researchers as a direct starting point for adaptation and advancement **DETAILS**. We will actively engage with the opportunity to learn more about the best ways to disseminate our work and work with other researchers to figure out if alternative plans would be more appropriate. We note that this plan is a living document.

[1] Anders Logg, Kent-Andre Mardal, and Garth Wells. *Automated solution of differential equations by the finite element method: The FEniCS book*, volume 84. Springer Science & Business Media, 2012.

[2] Saeed Mohammadzadeh and Emma Lejeune. Predicting mechanically driven full-field quantities of interest with deep learning-based metamodels. *Extreme Mechanics Letters*, 50:101566, 2022.